5-15-2019

HOW DO ENTREPRENEURIAL FIRMS APPROPRIATE VALUE IN BIO DATA INFRASTRUCTURES: AN EXPLORATORY QUALITATIVE STUDY

Hannes Rothe
Freie Universität Berlin, hannes.rothe@fu-berlin.de

Sirkka L. Jarvenpaa
University of Texas at Austin, sirkka.jarvenpaa@mccombs.utexas.edu

Anna Auguste Penninger
Freie Universität Berlin, anna.a.penninger@gmail.com

HOW DO ENTREPRENEURIAL FIRMS APPROPRIATE VALUE IN BIO DATA INFRASTRUCTURES: AN EXPLORATORY QUALITATIVE STUDY

Research paper

Rothe, Hannes, Department of Information Systems, Freie Universität Berlin, Berlin, Germany, hannes.rothe@fu-berlin.de

Jarvenpaa, Sirkka L., McCombs School of Business, University of Texas at Austin, Austin, Texas, USA, Sirkka.Jarvenpaa@mccombs.utexas.edu

Penninger, Anna A., Department of Information Systems, Freie Universität Berlin, Berlin, Germany, anna.a.penninger@gmail.com

Abstract

Recent technological advances such as in genome sequencing have exploded bio data infrastructures including those that comprise of generic - anonymized or pseudonymized - data. As open data, the bio data infrastructures do not constrain the final application context for their data. Rather it is up to complementors, taking the role of digital entrepreneurs, to appropriate value from this data through their revenue streams while at the same time scaling their operations and ventures. We undertake a qualitative explorative study of bio data ventures examining the tension of applying open generic genome data to specific contexts for customers while being able to scale their businesses. The study uses primary data from 26 interviews and secondary data to reveal six strategies that complementors use for value appropriation. We derive three mechanisms of appropriating value at different stages of the value chain for bio data analysis on open data infrastructures: data contextualizing, data decontextualizing, and data recontextualizing. The study sheds light to how bio data – which has received limited attention in information systems research – can be an important source of value appropriation in digital ecosystems.

Keywords: Value appropriation, complementor, data infrastructure, digital entrepreneurship, bio data

1 Introduction

With the rise of digital infrastructures and platforms, the question of how value is created and appropriated by complementors in digital ecosystems becomes ever more important (Constantinides, Henfridsson, & Parker, 2018; de Reuver, Sørensen, & Basole, 2017). Advances in digital technologies have led to increasing efficiencies of data transfer, storage, and processing enabling scalable architectures including platforms and infrastructures across a variety of domains (Sanchez & Mahoney, 1996; Tiwana, 2008, 2014). Data infrastructures can provide massive amounts of data to complementors that set up digital ventures to exploit data in real-time and around the globe (P. C. Evans & Basole, 2016; P. C. Evans & Gawer, 2016). Complementors integrate this data as boundary objects into their daily business practices and operations in order to create added value for their customers (Eaton, Elaluf-Calderwood, Sorensen, & Yoo, 2015; Ghazawneh & Henfridsson, 2013) as well as to appropriate value from it (Ceccagnoli, Forman, Huang, & Wu, 2011; Parker, Alstyne, & Jiang, 2017; Tiwana, 2014).

Yet, rarely has there been a dedicated focus on data itself (Cennamo & Santalo, 2013; McIntyre & Srinivasan, 2017; Zhu & Iansiti, 2012) and what data appropriation strategies are applied by complementors in the digital ecosystems literature. We refer to appropriability as the complementor's effectiveness in exploiting data by processing and translating it into a solution or service to a customer (Teece, 1986). Complementors apply data to the specific use cases of multiple target groups and finally grow the user base of infrastructures (Cennamo & Santalo, 2013; Iansiti & Zhu, 2007; Rietveld & Eggers, 2018). They solve concrete problems as they apply this data with IT artefacts in specific contexts of interest to their customers (Abbasi et al., 2016; Adomavicius & Tuzhilin, 2005). As such, complementors behave as digital entrepreneurs who use data in novel ways, create new opportunities in the market and help "moving the economy closer to the technological frontier" (Sussan & Acs, 2017). They utilize data to engage in future-shaping practices (Kelestyn & Henfridsson, 2014; OECD, 2015). At the same time, digital entrepreneurs need to find a sustainable appropriation strategy that is able to scale (Kelestyn & Henfridsson, 2014; Nambisan, 2016; Nambisan, Lyytinen, Majchrzak, & Song, 2017). This may inflict a tension for complementors – appropriate value from the data infrastructure within a specific customer context while simultaneously being able to scale operations and revenue.

Understanding this tension in the healthcare and bio technology space is timely. Particularly genomic data has exploded since the completion of the human genome project in 2003 (e.g., Lander, 2011). Advances in genomic sequencing technologies have led to massive cost reductions for producing genomic data. The numbers of ventures that make use of human, animal, microbiome, or plant genomic data is ever more growing (Birney, Vamathevan, & Goodhand, 2017; Jagadish et al., 2014). However, little research in information systems has focused on bio data and complementors leveraging bio data. Most genomics data is offered on open data infrastructures on a (pseudo-)anonymized basis. Data is offered as a common-pool resource to overcome the legal challenges and technical control associated with it (Koutroumpis & Leiponen, 2013). While more and more ventures are created, the commercial use of bio data has been largely overlooked in the domain of information systems (for exceptions, see e.g., Jarvenpaa & Markus, 2018; Vassilakopoulou, Skorve, & Aanestad, 2016). The rare writings on bio data have focused on infrastructure providers and their governance models (Jarvenpaa & Markus, 2018; Vassilakopoulou et al., 2016; West, 2017). For researchers and practitioners, it remains unclear how complementors on these open data bio infrastructures create value from data and appropriate financial value for their ventures. Hence, we raise the question: *How do complementors appropriate value from open bio data in digital ecosystems?*

We shed light on how complementors appropriate value from open data by outlining how they access, process, and transfer this data along the value chain. For this purpose, we introduce a qualitative exploratory study in the specific domain of bio data, namely genomics. Based on 26 interviews across thirteen complementors, investors and infrastructure providers that we complement with secondary data, we propose six value appropriation strategies for complementors in open bio data infrastructures. Finally, we infer three mechanisms of how financial value is appropriated from the data infrastructure: data decontextualizing, data recontextualizing, and data contextualizing.

2 Appropriation in open data infrastructures

Data is a key resource that ecosystems offer on top of infrastructures. However, providers of these infrastructures lack the ability to foresee all possible areas of application of their data (Tiwana, Konsynski, & Bush, 2010). Data-intensive businesses (new ventures and incumbents) serve as complementors to these infrastructures. Complementors can apply their knowledge about a specific problem space to novel use cases (Yoo, Henfridsson, & Lyytinen, 2010). Through these use cases, complementors can develop an array of functions and offer them to a variety of specific end-user needs (D. S. Evans, Hagiu, & Schmalensee, 2008; Ghazawneh & Henfridsson, 2011; Huang, Ceccagnoli, Forman, & Wu, 2009).

We conceptualize complementors that utilize data from digital infrastructures to create value for customers as digital entrepreneurs. The generative nature of these infrastructures allow entrepreneurs to engage in future-shaping practices (Kelestyn & Henfridsson, 2014; OECD, 2015, Dörbecker, Böhm, & Böhmann, 2015). Complementors not only need to create but also capture value (Teece, 1998). But besides appropriating value from specific customers, complementors need to scale their activities of value creation, delivery, and appropriation (Kelestyn & Henfridsson, 2014). This tension of appropriation and scaling is particularly evident if external venture capital is involved (Autio, Nambisan, Thomas, & Wright, 2018; Nambisan & Baron, 2013).

To appropriate financial value from open data, complementors need to have an ability to *protect their own resources* and operate on the *data infrastructure* (Pisano & Teece, 2007). When it comes to the ***data infrastructure***, complementors have to have downstream capabilities that sustain their ventures particularly in terms of sales as sales is critical for securing future external investment (Ceccagnoli, Forman, Huang, & Wu, 2012). The strategy literature has long argued that downstream capabilities, including manufacturing, marketing and sales, are essential for creating a value chain (Teece, 1986). At the same time the necessary capabilities for performing these activities successfully are hard to attain, and may constitute competitive advantages (e.g., Barney, 1991) to complementors (vis a vis infrastructure providers) in digital ecosystems. While these complementors innovate on data resources and differentiate among each other, the infrastructure provider can become a competitor to these complementors and expand its reach across more and more customer segments. In fact, from the provider's perspective, the complementors' superior capabilities in creating value from data is the primary reason for opening data resources in the first place (Benlian, Hilkert, & Hess, 2015). At the same time, making data resources openly available invites competition between complementors. The generative functions and boundary resources provided on the infrastructures eases an early set-up of value creating activities by new entrant complementors as well as supports its scaling.

It remains an open question to what extent the general logics of collaboration and competition in digital ecosystems apply to open data – particularly in the domain of bio data. Genome data remains fragmented, non-standardized, and much of it indeterminate and hence error prone in open data infrastructures (Vassilakopoulou et al., 2016). The data quality issues are exacerbated by the multidimensionality of the data and by rapidly advancing scientific insights (Lee, 2015).

In order to understand the appropriation of value from such data, we need to gain an understanding about the value chain of data. Much of the value resides in data analysis. Data mining literature provides multiple procedural models on how knowledge is extracted from data, and finally how value is created. Among them, the knowledge discovery in databases (KDD) model is most prominent and serves as a good baseline for understanding how single or multiple parties (1) collect & select, (2) integrate, preprocess & clean, (3) transform, (4) mine, and finally (5) interpret data throughout commercial and research domains (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Hartmann, Zaki, Feldmann, & Neely, 2014). It can be applied as a value chain for the biotech industry (Jagadish et al., 2014). With regards to bio data this procedure describes the "first mile of data generation to the last mile of using data with customers" (interview #3). This entails (1) collecting & sequencing blood or tissue samples, (2) preprocessing the data, (3) discovering relationship models to predict a disease, perform genotyping, or find associations between phenotypes and gene mutations, and (4) interpreting the results, for instance to infer therapeutic advise (Diao, Roy, & Bloom, 2015; Project MinE, 2015).

There is a wide range of open data infrastructures in the domain of genomics and bio data. The European Molecular Biology Laboratory, among the largest infrastructure providers for bio data from research, names 724 web services to access and incorporate bio data themselves, for instance on gene expressions, protein sequences, literature, ontologies, and the genome (McWilliam et al., 2013). Data and services on these infrastructures can be used to build complex data pipelines. Raw bio data is regularly large in storage size. Whole genome sequences – bits of long strings of data – may in sum range from 15 GB to 15 TB size depending on the resolution and technology applied (Voelkerding, Dames, & Durtschi, 2009). Physical tissue samples are oftentimes stored in bio banks in order to be sequenced and transformed into data therafter. They are accessible for research projects. However, it is not just the size of bio data but also legal considerations that matter. Working with bio data involves many requirements around informed consent of the individual. Regulations such as GDPR limit future usage of collected data to the purpose as described in the consent. Raw bio data is regularly pseudonymized, anonymized, or aggregated in order to be saved, shared, and later processed without having to go back to gain additional consent from individuals. Gene data, for instance, is stored as variance alongside a reference gene model, such as the human reference genome. At the current time, no global data standards exist for genome data although such standards are under development in the Global Alliance for Genome and Health Data (Birney et al., 2017).

Protection regimes of complementors block infrastructure providers and new entrants from exceeding the complementors' downstream capabilities (Ceccagnoli et al., 2011). In other words, intellectual property rights associated to activities along the downstream set up entry barriers to incumbents and other new entrants, and allow complementors to derive additional financial value (Arora, Fosfuri, & Gambardella, 2001; Ceccagnoli et al., 2012; Oxley, 1999). While activities, including techniques, procedures or algorithms, can be protected to various degree, data provided on the infrastructure largely remains a common-pool resource. Unless bio data is kept entirely secret, it is an open resource. Intellectual protection regimes, such as copyright or database right, weaken over time. In contrast, high transaction costs may serve the purpose of protecting bio data. Transaction costs are associated to the necessary efforts taken for access, integration and quality verification of bio data. These are particularly influenced by the heterogeneity of bio data in terms of their formats and content.

Koutroumpis and Leiponen (2013) underscore the importance of meta data when it comes to handling the heterogeneity of data. Metadata might include its origin, the characteristics, history, legal status, consent, and furthermore characteristics, necessary to describe and apply the context of data. In the case of genome data, metadata also includes the annotations conveying "the location and function of genes" (Lee, 2015), which might be structured or unstructured data. Contextual information is critical, for example to develop generative algorithms and derive predictions (Adomavicius & Tuzhilin, 2005). Explicitly adding contextual information to a core message "elucidate the situation in which the message was created" (Te'eni, 2001, p. 266). Contextualization comprises additional information about the origin of the message, its originators, and about the message itself. Senders of a message add contextual information where they do not assume that a recipient knows it already (Horton & Keysar, 1996). Senders weigh costs and benefits of providing contextual information (Katz & Te'eni, 2007). Following this logic, contextual information becomes particularly important under uncertainty, where perspectives of information providers – here data providers – and receivers – complementors and end-users – might differ (Katz & Te'eni, 2007). Misunderstandings are likely to occur (Weick & Meader, 1993), where activities are non-routine (Majchrzak, Malhotra, & John, 2005).

In their data-based products and services, complementors leveraging data infrastructures along the value chain add, but also reduce meta data from their customers. Data decontextualization may enable a seamless experience of their products and services (Yu & Woodard, 2008). However, adding meta data is critical in generative activities as it forms the basis of data contextualization or data recontextualization. Thus, while contextualization of data is critical to create and appropriate value, so can be data decontextualization.

3 Research Methods

Given our lack of understanding how value is appropriated from open data in data infrastructures, we introduce a qualitative-explorative field study (Klein & Myers, 1999; Walsham, 2006). Field study research is interpretative in nature, theoretical abstraction and generalization is carefully related to the details of the information collected in the field (see also Gregor, 2006). This study helps us understand what type of challenges complementors face as digital entrepreneurs that appropriate value from bio data infrastructures, and which strategies they apply to overcome them.

id	Institution type	Role	id	Institution type	Role
#1 (2x)	Research Institute A / Data platform provider (>100 PB data)	Partnership manager	**#13** (2x)	Venture capital (focus on BioTech & Health)	Manager Healthcare investment
#2	Research infrastructure provider A	Industry partnership manager	**#14**	Bio data venture G	Chief Executive Officer
#3	Research infrastructure provider A	Research partnership manager	**#15**	Bio data venture H	Chief Executive Officer
#4	Bio data venture A	Product manager	**#16**	Bio data venture I	Chief Executive Officer
#5	Research Institute A	Innovation manager	**#17**	Bio data venture J	Chief Executive Officer
#6	Bio data venture B	Chief Executive Officer	**#18**	Biology department in research institute B	Impact facilitator
#7	Bio data venture C	Chief Science Officer	**#19**	Private physician practice	Physician
#8	Research infrastructure provider A	Partnership manager	**#20**	Bio data venture A	Chief Executive Officer
#9	Research Institute B	Business development manager	**#21**	Research institute C	Head of research infrastructure
#10	Bio data venture D	Chief Executive Officer	**#22**	Bio data venture K	Chief Executive Officer
#11	Bio data venture E	Chief Science Officer	**#23**	Bio data venture L	Chief Executive Officer
#12	Bio data venture F	Chief Executive Officer	**#24**	Bio data venture M	Chief Executive Officer

Table 1. Interview data

We selected our interview partners from complementors in the domain of genomics that have a steady stream of revenue from their core products or services and have attracted external investment backing. Starting with a convenient sampling, we contacted additional complementors until we reached closure (compare Eisenhardt, 1989). Data providers, a biotech venture capital investor, and a physician were also included to triangulate statements of the complementors (see Table 1). Given possible influences from environmental factors, particularly from regulation on storing, sharing, and utilizing genome data (i.e., GDPR) the sample was limited to European complementors leveraging open data infrastructures.

Each interview lasted between 45 and 90 minutes. From 26 interviews (see Table 1) all but five were fully transcribed, the rest were documented in field notes. We employed four guiding questions, and left freedom for active conversation from which further, unexpected information might arise (Silverman, 2016):

0. *Give me an overview about the history of [your company name]*. Introduce me to your role and responsibilities. How are you funded and what is your aim for your next funding?
1. *For whom and how do you create revenue*? Along which customer channels?
2. *Who are your main partners and what resources do they provide?* What data is provided by others? (raw data, aggregated models) How is this data used within your company? Are third parties involved in data collection, preparation or analysis, and if so, how (i.e., algorithms, development)?
3. *What is the output that you provide to your customers?* How diverse are the fields of application for your product/service (within/across locations, industries …)? What is your output format and are there client specifications for your output?

In addition to this conversational data, we triangulated the information about each use case emerging from the data collection by adding information from press releases about the ventures, their product catalogues and websites, information from venture databases such as Crunchbase.com and Angellist, as well as from the company database Amadeus.

To analyze and interpret the data, we (1) prepared case reports for each complementor. Case reports consisted of each complementor's meta data (i.e., size, age, investment, data sources), summarized core operation of the core products and services, explained how value is created along the value chain, outlined the customer segments, described input data and output formats, and briefly described activities with main partners. For this purpose, quotes from interview data were extracted into the case report and information from secondary data was added. Afterwards, (2) we sorted the complementor's products & services into stages along the value chain as we iterated between raw data and conceptual knowledge (Klein & Myers, 1999). These case reports were the basis for comparing value appropriation between complementors. For each case, we discussed its activities along the value chain and found where the (3) core value was appropriated, informed by the complementor's main source of revenue as well as its current focus of their growth strategy. We discerned between activities (3a) that are employed by the complementor itself and created an output on a particular stage of the value chain (i.e., selecting raw sequence data), (3b) activities performed by a third party (i.e., outsourcing DNA sequencing to another laboratory), as well as (3c) activities where the complementor may not be active on a stage along the value chain itself, but offers an infrastructure to another partner without appropriating its value (i.e., offering a data infrastructure with an interface to upload DNA data may serve as collecting raw data, however the infrastructure mainly appropriates value from sending a final report to physicians). Afterwards, (4) we compared all strategies of value appropriation across cases and along the value chain and found consensus among the author team on a pattern of six strategies for appropriating value. This result was discussed with a domain expert in a two-hour meeting. While each strategy offers information on the practice of how data is processed and how revenue is created, we (5) revisited each strategy in order to evaluate scalability of complementor's activities alongside current revenue streams. This evaluation was based on a mapping to what extent complementors standardize input and output data formats across customers.

4 Findings

Based on the inductive analysis of field data, we present six strategies of complementors to appropriate value from open data infrastructures in the field of genomics. Each strategy is briefly summarized and illustrates typical customer segments. In order to share information about the complementor's role in the data infrastructure, we pay particular attention to input and output formats and position the complementor along the value chain. Finally, we reflect the role of meta data and annotation for each strategy. We conclude by relating the strategy in terms of data contextualizing, decontextualizing, and recontextualizing and how open and proprietary data influence the ability to appropriate value. The standardization of input and output formats has an important role in terms of the complementors' ability to scale business activities and revenue streams.

Appropriation Strategies	**Value Chain for Data Analysis**				
	Data Selection	**Data Integration & Preprocessing**	**Data Transformation**	**Data Mining**	**Data Interpretation**
Custom Full Service (A)					
High Quality Data Service (B)					
Vocabulary Generation (C)					
Configurable Model (D)					
Care Service (E)					
Workflow as a Service (F)					

Table 2 Appropriation strategies on the value chain. core activity (dark), infrastructure (light)

Table 2 provides an overview of the findings in terms of appropriation strategies and how strategies map onto the value chain. We discern between core activities applied as part of the appropriation strategy (dark field) and infrastructures provided to customers or partners from which value is not necessarily appropriated (light grey field).

Custom Full Service Strategy (A) appropriates value by offering an integrative service across the entire value chain.

Customer segments: This strategy applies to customers ranging from the support of biological discovery to diagnostic or drug discovery, effectively addressing various markets, amongst others pharmaceuticals, agriculture, food production, and contract research organizations (CRO).

Position along the value chain: “Customers buy our expertise in combination with the results of our analysis” (#22). They outsource their work to complementors leveraging the data infrastructure. The complementor’s position involves activities along the entire value chain – from data selection, preprocessing, and analyzing bio data to providing concrete interpretations. At the same time, this strategy helps customers to avoid potential backtracking of their data sources and data utilization by their competitors; not leaving traces that give competitive information about their activities, for instance on drug discovery for pharmaceuticals.

Appropriation of value: This strategy was deployed by many complementors early on but abandoned later as it "doesn't scale. Effectively, you build [the service] every time hand in hand with a customer" (#23). For this reason, it deems only viable for high margin projects, which regularly does not apply to "academic customers, because research groups send inquiries but show no willingness to pay" (#22). The strategy is limited by the quality of open data that often was not collected with secondary use in mind. For example, the meta data on consent may be missing or lacking which leaves legal uses ambiguous. However, it enables the creation of initial revenues and helped many complementors to form a venture in the first place. At the same time, this strategy opens the chance to learn about potentially scalable models. For instance, "the prediction of grain yield on a field […] we learned about the relationship between phenotype and genotype and once we created a model, we are able to predict for millions of cross breedings which have the highest potentials on a particular location." (#22). Due to the deep integration of the customer, however, this customer might determine rights of utilizing such models. In one case, the venture got stuck on contracts with a large pharmaceutical company. Even though it developed a potentially scalable model and developed the software, contractual agreements prohibited its outside use. In other cases, pharma companies write contracts that demand that all meta data and annotations belong to it or derivatives from it cannot be used freely outside of the relationship. Thus, Full Service Strategy does not always allow complementors to hold onto their developed intellectual property. This may decrease the chance to scale, as complementors can only transfer their experiences but have to start anew each time.

Interaction with the data infrastructure (input/output): For complementors, this strategy means broad exposure to public and proprietary data in a variety of contexts as well as accumulation of experience along and across the entire value chain. They choose input and output data individually for each customer and highlight the relevance of interpretability and understanding in particular contexts of the application. "We don't believe that providing this directly to the consumer [customer of the complementor] is actually going to help them. We really think that actually is going to confuse that because you have 30% chance of likelihood of having prostate cancer. What you do? What does that mean?" (#11). Instead, these complementors select the data to be presented as an output of their service precisely to the need of their customers, as they "give people a glass of water instead of leaving them soaking wet" (#23). This strategy offers full flexibility to select and adapt any kind of input data to projects over a wide set of customer requirements. The appropriated value stems from outsourcing large parts or the entire data pipeline from customers to complementors which "enable[s] the customers to activities they are not able to perform on their own” (#22). Output formats are tailor-made to the customer needs. In summary, value is appropriated by ***data contextualization.***

High Quality Data Service Strategy (B) appropriates value from creating a proprietary database with additional meta data (e.g., personal information, health data) that accompany open data.

Customer segments: On one side, raw genome data can be accessed by research groups or contract research organizations that follow downstream the value chain. On the other, consumers with genetic disease (e.g., rare diseases) upload their data or send samples to be processed by another third party.

Activity along the value chain: Complementors operate at the beginning of the value chain as they select data for their purposes. They build up proprietary data bases of bio data, which become their main asset and is passed on downstream. While the data is largely proprietary, these complementors rely on public data for quality checks and triangulation. “So as you can imagine genetic data follows particular kinds of structures. You can spot, if someone has altered it. It’s probably possible to alter it in such a way that it cannot be spotted – but we have a lot of checks in place to ensure none of the obvious […] We check if its already been available online or if it’s suspiciously close to anything that is available online. […] if you wanted to fabricate a genome and want to be really careful with it you’d want to structure in ways that aren’t normal that is not what you’d expect from a normal person who inherited their DNA blocks from someone else” (#6). In summary, the complementor provides its own infrastructure to upload raw bio data, does initial preprocessing to check the data quality with the help of open data, and passes it downstream the value chain.

Appropriation of value: Customers pay for getting access to raw data from individuals who have given their explicit consent to share it. Donors of bio data, including DNA or RNA, share their data regularly for supporting a cause, for instance research on a precondition or disease. Besides this altruistic motivation, during our interviews, we saw a shared revenue model being tested by our interviewee and another being mentioned, where CROs pay a fee to donors and the complementary serves as an intermediary. The complementor finally decides on the access to their proprietary data set.

Interaction with the data infrastructure (input/output): The complementor uses a predefined meta model to represent sequenced DNA data on their infrastructure, and ask suppliers either for a pre-defined data set – such as already sequenced genome from 23andme – or a cell sample, i.e., blood or saliva, to get it sequenced by a third party into the same predefined format.This metadata includes information on consent, further biological information such as the structure of a DNA strand in addition to data on the human genome string, or personal contact information to allow longitudinal analysis. The latter may allow "to go back to this individual and ask a follow up questions or a specific hypothesis how do individuals with this particular genetic variants fare differently” (#6). Complementors define which services can be integrated into their infrastructure. Independent from individual customer, they pre-define the input data formats and what kind of metadata is stored. Hence, contextual information in forms of metadata and annotations is stripped from the input data and a standardized dataset is offered downstream. Thus, we speak of ***data decontextualization***.

Model Enhancement Strategy (C) enhances existing models, such as vocabularies, ontologies, or reference models, that can be utilized by their customers. Models are refined through manual coding, triangulation with open and proprietary datasets, and machine learning on new structured and unstructured data instances.

Customer segments: Customers potentially span across healthcare industry’s upstream and downstream activities. We found many customers are in upstream healthcare domain, including pharmaceutical companies in early stage discovery as well as downstream healthcare domain such as clinical trials and clinical care. This strategy is also used along with other strategies particularly with a Configurable Model Strategy or a Care Service Strategy (to be discussed below).

Activity along the value chain: Enhancing reference models or ontologies involves specialized domain knowledge that is, for example, applied to integrate, pre-process and transform data into a usable model for downstream activities. It involves "both the machine learning and human efforts to enrich those. So if you look at something like Human Phenotype Ontology (HPO) that has about 17,000 entries, I think. There’s less than one synonym per entry. You got heart defects, that could also be defective hearts, defects of the heart, cardiac defects, all those different terms, but then in the HPO, there’s less than one synonym per entry. That’s no way near the amount [our customers] need." (#7)

Appropriation of value: These proprietary vocabularies or reference models are the core resource that complementors use to appropriate value. The refined model remains hidden; as one founder stated, "it will all be a black box. All the information and all of the algorithms will be in a closed system that we maintain ourselves.” (#12) Hence, they are not sold directly. Instead, models can be accessed from proprietary cloud environments via application programming interfaces (APIs). They are integrated into customer software or web-services.

Interaction with the data infrastructure (input/output): Healthcare clinical practice has particularly high-quality standards for data. Often the open data from data infrastructures originates from publicly funded research projects and such research data does not meet clinical standards. "What [public bio database] filters down in the academic domain does not suffice for our customers [in pharma and the processes are] refined by us to make them fit for use in commercial products and industrial processes" (#17). For instance, while specific reference models on open data infrastructures are not good enough for industry application, one complementor developed new reference models by triangulating existing information using machine learning. Another designed new ontologies, "lists of things, list of diseases, list of genes, list of tissues, whatever. Those are predominantly open data; those are predominantly from the open source world which is brilliant because that’s what we want to be able to connect things into." (#7) Altogether, these ontologies help integrating data from different sources, and they can be used to predict the context of a new data instance – i.e., solving the semantics problem of synonyms, homonyms, and antonyms. Given this ontology, for instance, an algorithm is able to assess whether 'hedgehog' refers to a spiky animal or the signaling pathway which transmits information to embryonic cells. The models are used downstream the value chain to link or compare new data instances. These models consider and partly predict contextual information to derive recommendations (Adomavicius & Tuzhilin, 2005). Initially, the models that are being enhanced were created through generalization from raw data. Metadata – such as the exact location of a cell from which the DNA strain was extracted, patient information, or the context in which the data was applied– was stripped off in order to construct a generalizable model. Customers of these enhanced models pay to impute missing metadata from indeterminate open data. Hence we find appropriation from ***data recontextualization***.

Configurable Model Strategy (D) integrates public and proprietary data sources and make them available over a web interface.

Customer segments: Customers range from academic research groups over contract research organizations (CROs) to large pharmaceutical companies.

Activity along the value chain: Complementors integrate data from different infrastructures and transform them into a ready-made form for customers to perform ad-hoc analysis. They develop a search functionality and methods of data integration that are scalable towards new data sets. Therefore, they employ public or private ontologies, i.e., the results of complementors following a Model Enhancement Strategy. For public data sets alone, research infrastructure providers name more than 30 key data infrastructures, ventures with a configurable model strategy published lists between 60 to 134 bio data infrastructures. In effect, these complementors solve the issue that “you have very fragmented information [...] and those companies combine the information” (#2). In addition, complementors provide a user interface that is easy-to-use for practitioners in the field and allows addition of metadata by customers, for instance saving the search query, adding comments to datasets, links to publications, information on sequencing technology, the tissue etc. One complementor exemplifies this, "We don't know anything about the patient we only know something about the variant [comparison of DNA with a reference model]. It's a bit like going to Google. […] That person who makes a query knows a lot more about the patient. They might actually add more information." (#15).

Appropriation of value: Financial value is appropriated by adding proprietary datasets from companies and use the search across the public and proprietary dataset. Additionally, these complementors offer a web interface to raise the ease of using open data, which is regularly difficult and cumbersome for professionals. These complementors unanimously provide a freemium model, where searching public databases is free. Premium functions include applying the superior search function to databases from customers themselves, licenses to other proprietary datasets, or complementary functions such as the ability that "you can go back [to your search query], you can put those alerts. The fact that you can go back to it yourself and redo the search in an easy way, that you can share this. It's all about sharing." (#24) In the future, added metadata such as the history of search queries may establish a valuable resource on its own.

Interaction with the data infrastructure: At the beginning, open data served the role to train and test the system with customers and build a product that is able to attract external investments. Results are accessed over a web- or app-based user interface that is standardized across customers. However, integrating open data with proprietary data for paying customers (i.e., comparing individual data from patients or end-users with biomarkers) enables real-time comparison of different bio datasets. Hence, by allowing triangulation and manually adding metadata, such as annotations, links to datasets or publications, this strategy also appropriates value from ***data recontextualization***.

Care Service Strategy (E) provides tailored ready-made reports on a range of topics, including paternity, maternity, sibling, ancestry or ethnicity testing, or human and pet health.

Customer segments: Complementors provide reports to physicians (business-to-business) or consumers directly (business-to-consumer). Health reports mostly target physicians. They involve diagnostics and inform therapeutic recommendations as they reveal disorders or predispositions. Even though such health reports are highly prescriptive - sometimes not only including a description of the gene mutation but also concrete therapeutic advises – it is up to a physician to communicate a diagnosis to patients and make final decisions on treatments.

Activity along the value chain: Complementors mainly mine and interpret the data on the basis of models developed and evaluated in public research. Their reports largely consist of information on the relationship between data points, such as genome mutations in comparison to a reference model. These complementors look at the "variation in your DNA compared to the healthy population you will basically get the disease […] and the more variation [in this context] means the more likely that you get the disease" (#14). In effect, this appropriation strategy regularly applies to areas where there is well-researched and already trained predictive model that can be used on a new data instance to derive risk profiles. For example, the complementor predicts probabilities based on a panel of "genes that are specifically related to a particular condition: diabetes I, diabetes II, hypertension, heart disease” (#11). These panels either stem from published research accessible through open data infrastructures such as Pubmed or they are trained on a mixture of publicly available and proprietary data. For ventures that follow a care service strategy, the "bottleneck is the management processing, analyzing and interpretation of the data.” (#11).

Appropriation of value: Revenue stems from one-time payments from physicians and consumers for tailored reports. When it comes to health reports, physicians finally evaluate, interpret and draw conclusions. It remains "a human judgement process as the outcome of the test; it is a combination of using different software and then making a human interpretation on the result" (#13). Having a medical license provides the right to diagnose, thus including a physician into the process might be a legal requirement. The strategy is highly scalable across new data instances and new disease domains as long as sufficient research exists to develop robust predictive models. How operations are applied to data to generate the report remains a trade secret. Besides different customer segments, the appropriation strategy differs distinctively from the Custom Full Service Strategy (A). This latter strategy suffers from the lack of interpretability of bio data reports. Complementors that follow a Care Service strategy either do not perceive similar lack of interpretability of their reports when they target physicians or they provide training and guidance to help customers (physicians or consumers) understand their output data.

Interaction with the data infrastructure: Each new data point is collected from patients by physicians or consumers either via (blood, feces, saliva) samples and sent to third parties for sequencing or involves access to already sampled data from other complementors, such as 23andme or Ancestry.com. In contrast to complementors that follow a Model Enrichment strategy, complementors deal with a perceived lack of data quality from open data infrastructures by limiting their influence on the value creation process, i.e., “we are very concerned with the data quality of Clinvar - I think it is only used as a benchmark or in some visualization tool from us” (#13). Hence, a validated and fully trained model, i.e., a knowledge relationship between biomarker and a disease, is used for a new instance. Input and output data is predefined by the complementor. In order to enable a seamless experience of their products and services, and comparable to the High Quality Data Provider strategy, these complementors, also select and hide meta-data and annotations from their end-users (i.e., Yu & Woodard, 2008) to keep control of the models and retain their intellectual properties. Also, complementors only provide key information to the customer which is easy to understand and helps guiding actions. Thus, while such complementors consume highly contextual information (i.e., raw DNA sequences or samples), they only share limited meta-data with their customers. Limited revealing of meta-data impedes imitation. They appropriate value through ***data decontextualization***.

Workflow as a Service Strategy (F) enables customers to standardize data and connect databases and services in the data infrastructure into a coherent pipeline along the value chain.

Customer segments: Customers of these complementors are in pharmaceutical, cosmetic, and agritech industries.

Activity along the value chain: The complementor offers a modular infrastructure to support all activities along the value chain; “We are in the business of deploying software enabling these people to get as much out of the data as possible. […] workflow management system to scale up large scale data analysis [and add] the part of the infrastructure that does the data cataloguing. Where is my data? For what purpose is it used? What experiment? Who has access to it?" (#10). Complementors enable other actors in the data infrastructure to combine multiple services into a coherent data pipeline, essentially building on "an ever-growing library that can be combined to create complicated workflows and support analysis in genomics, metagenomics, and visualization.".

Appropriation of value: The value stems from outsourcing large parts or the entire data pipeline from customers to complementors which "enable[s] the customers to activities they are not able to perform on their own” (#22). Customers pay for data management services. Complementors do not create revenue on data, instead they are "a software platform to handle data, either open or private data. […] The data we are getting from the customer is their data. Whenever the customer wants to finish a contract they have access, copy it back, and we close and delete everything." (#10).

Interaction with the data infrastructure: Complementors predefine which databases and services can be accessed on their infrastructure and customers decide freely what experiments they want to pursue or outputs they seek. "What we try to do on the [infrastructure] is to represent the knowledge base; the baseline for a specific context of a customer so he knows how is data adding value to the common good." (#10) Hence, this strategy appropriates value through ***data recontextualization***.

5 Discussion

Our study is one of the few studies in information systems research that explores bio data as the source of appropriation of value in a digital ecosystem. The complexity of data and its analysis in the field of genomics and bio data creates a broad opportunity space for numerous complementors to undertake specific tasks and operations on data along the value chain (see also Table 3). This study provides insights into how this complexity is appropriated by complementors. Following up on six strategies for appropriating value on the open data infrastructures of genomics, we find that each strategy differed in how it opens to the data infrastructure, either using pre-defined or tailor-made input and output formats. Additionally, they offer different protection regimes. These regimes open varying opportunities to scale. We derived three mechanisms for value appropriation from open bio data: data contextualization, data recontextualization, and data decontextualization. Figure 1 summarizes these arguments.

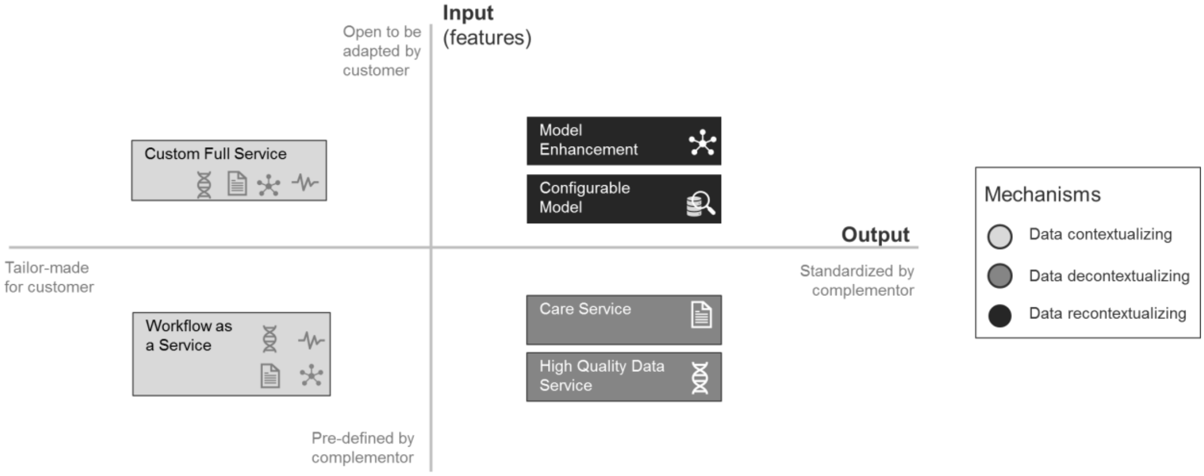

Figure 1 *Mechanisms of bio data appropriation*

Appropriation by contextualization: We find that a Custom Full Service strategy (A) as well as the Workflow as a Service Strategy (F) enable customers of the complementors to define their customized individual output. Both strategies span the entire value chain, but differ in the extent of their support services as well as intellectual property. When complementors focus on managing workflows, data remains proprietary to the customers. At the same time the methods integrated as modules on the workflow management system remain proprietary to the third party but may generate economies of scale. On the other hand, customers of complementors with Full Service strategy outsource one-time analysis or the development of tools and applications. Using this strategy comes with the risks to be bound to one specific context or even one large partner, i.e., a corporation or university. Oftentimes, this customer owns the intellectual property. As we see, complementors with a Workflow as a Service Strategy may rely on methods of other third parties and also fail to build up a protection regime themselves. We learned that a Custom Full Service (A) strategy was initially used by a large number of complementors. However, most complementors adapted their appropriation strategy when they found out that either the fields of data application or the practices of how data are processed were not scalable. As one interviewee pointed out "when you do analysis you do one report, you cannot copy and paste one report from one customer to another one. So you have to do almost a custom work […] and you can only scale it by more contracts and people" (#10).

Appropriation by decontextualization: Both strategies, the High-Quality Data Service (B) as well as the Care Service Strategy (E), apply established models of data representation or data analysis to derive value for their customers. While both are at the opposite ends of the value chain, their input and their output data remain predefined by the complementor and cannot be adapted by either customer or suppliers. This gives complementors good opportunities for scaling and is a strategy that services to consumers directly – either as donors of data or as end-users for reports. While this opens opportunities for gaining revenues and growth, which the largest complementors in our sample also suggest, it is also becoming a highly saturated and highly competitive domain. Given our information from company databases Crunchbase and dnatestingchoice.com, the market for ancestry reports names more than 55 worldwide competitors, and health testing with more than 100.

Appropriation by recontextualization: Complementors that follow a vocabulary generation strategy (C) and the configural data model strategy (D) standardize their data output while they adapt data input to their customers. Both appropriation strategies link various individual datasets, add metadata to existing datasets, and make them again accessible to customers. For this purpose, they enhance and build upon generalized models and apply them to a specific context by adding this metadata. The results are presented in form of standardized APIs or result lists that allow for comparison between datasets and models on the complementor's web interface. Complementors who appropriate value through recontextualization serve in B2B markets in the middle of the value chain. For this purpose, products are highly specialized. Products remain black boxed to customers to some extent, but their main value comes from the creation of proprietary metadata and annotations. Given the necessity to solve a severe but generic issue of business customers, products regularly have an industry focus or serve an emergent field in genomics, e.g., epigenetics, plants, or microbiome.

			Applied appropriation strategy					
Ven-tures	**Size** (employees)	**Year of foundation** (pivot)	**Workflow as a service**	**Custom full ser-vice**	**High qua-lity data provider**	**Model enrich-ment**	**Configu-ral data model**	**Care service strategy**
D	25-50	2008 (2013)		before pivot				
E	1-10	2017						
I	1-10	2017						
K	25-50	2012						
L	25-50	2004						
B	1-10	2017						
J	1-10	2007						
C	25-50	2011 (2015)		before pivot				
A	25-50	2015						
H	1-10	2014						
M	25-50	2011						
F	25-50	2012 (2016)				before pivot		
G	150-250	2012						

Table 3. Appropriation strategies of complementors in the sample

5. Conclusive remarks

This inductive study explored value appropriation strategies of complementors in data infrastructures on bio data. Bio data, particularly genomic data, grows at an exponential rate. The largest proportion of data is collected in research or within national endeavors, and applied by private companies. Prior research on bio data has largely examined data infrastructures from the perspective of infrastructure providers (Jarvenpaa & Markus, 2018; Vassilakopoulou et al., 2016; West, 2017). This infrastructure perspective highlights that complementors appropriate value from generic boundary resources by applying them in concrete contexts (Eaton et al., 2015; Ghazawneh & Henfridsson, 2013).

Our research shows how complementors appropriate value along the value chain and questions the assumption that data contextualization enables value appropriation. We present six strategies from complementors and how they appropriate value within bio data infrastructures. We conceptualize complementors as digital entrepreneurs (Davidson & Vaast, 2010; Kelestyn & Henfridsson, 2014). While they innovate on open generative resources, they need to appropriate value not only in terms of generating revenue but also scale their business. While open data infrastructures do not provide data for particular contexts, complementors appropriate value along the value chain following three mechanisms: (1) data contextualizing involves a service along the entire value chain that helps initiate a venture, but may impede appropriation in the long run as it is hard to scale and may not support the creation of sustainable resource protection regimes; (2) data decontextualizing involves B2B / B2C services or products that apply established models and standardizes input and output data - either at the beginning or the end of the value chain. Intellectual property can be created and maintained as customers use this service as a black box. (3) Data recontextualizing involves the offer of highly specialized B2B services along the value chain. Complementors infer metadata and annotations from public and proprietary data, which by itself becomes a valuable resource. We also briefly discuss how complementors shift between data contextualization, data recontextualizing, and data decontextualizing. This exploratory study strives toward a typology of appropriation strategies in a poorly understood but societally strategic area. Given convenience sampling, our findings must be considered tentative.

Acknowledgement: We thank intergovernmental life science infrastructure ELIXIR for their support.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2).

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306–333. https://doi.org/10.1002/smj.821

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions On*, *17*(6), 734–749.

Arora, A., Fosfuri, A., & Gambardella, A. (2001). Markets for technology and their implications for corporate strategy. *Industrial and Corporate Change*, *10*(2), 419–451.

Autio, E., Nambisan, S., Thomas, L. D. W., & Wright, M. (2018). Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems. *Strategic Entrepreneurship Journal*, *12*(1), 72–95.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1).

Benlian, A., Hilkert, D., & Hess, T. (2015). How open is this platform? The meaning and measurement of platform openness from the complementors' perspective. *Journal of Information Technology*, *30*(3), 209–228.

Birney, E., Vamathevan, J., & Goodhand, P. (2017). Genomics in healthcare: GA4GH looks to 2022. *BioRxiv*, 203554.

Ceccagnoli, M., Forman, C., Huang, P., & Wu, D. J. (2011). Co-creation of value in a platform ecosystem: The case of enterprise software. *Management Information Systems Quarterly*, *36*(1), 263–290.

Ceccagnoli, M., Forman, C., Huang, P., & Wu, D. J. (2012). Cocreation of value in a platform ecosystem! The case of enterprise software. *MIS Quarterly*, 263–290.

Cennamo, C. (2016). Building the value of next-generation platforms: the paradox of diminishing returns. *Journal of Management*, 0149206316658350.

Cennamo, C., & Santalo, J. (2013). Platform competition: Strategic trade-offs in platform markets. *Strategic Management Journal*, *34*(11), 1331–1350. https://doi.org/10.1002/smj.2066

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4).

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and Infrastructures in the Digital Age. *Information Systems Research*, *29*(2), 381–400.

Cusumano, M. A., & Gawer, A. (2002). The Elements of Platform Leaderhsip. *MIT Sloan Management Review*, (Spring), 51–58. https://doi.org/10.1371/journal.pone.0015090

Davidson, E., & Vaast, E. (2010). Digital entrepreneurship and its sociomaterial enactment. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (pp. 1–10). IEEE.

de Reuver, M., Sørensen, C., & Basole, R. C. (2017, April 11). The digital platform: a research agenda. *Journal of Information Technology*, pp. 1–12. https://doi.org/10.1057/s41265-016-0033-3

Diao, Y., Roy, A., & Bloom, T. (2015). Building Highly-Optimized, Low-Latency Pipelines for Genomic Data Analysis. In *7th Biennial Conference on Innovative Data Systems Research*. Asilomar, California, USA.

Dörbecker, R., Böhm, D., & Böhmann, T. (2015). Measuring Modularity and Related Effects for Services, Products, Networks, and Software--A Comparative Literature Review and a Research Agenda for Service Modularity. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 1360–1369). IEEE.

Eaton, B., Elaluf-Calderwood, S., Sorensen, C., & Yoo, Y. (2015). Distributed tuning of boundary resources: the case of Apple's iOS service system. *Mis Quarterly*, *39*(1), 217–243.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.5465/AMR.1989.4308385

Eisenmann, T. R., Parker, G., & Van Alstyne, M. (2009). Opening platforms: how, when and why?

Platforms, Markets and Innovation, 131–162.
Evans, D. S., Hagiu, A., & Schmalensee, R. (2008). *Invisible engines: how software platforms drive innovation and transform industries*. MIT press.
Evans, P. C., & Basole, R. C. (2016). Revealing the API Ecosystem and Enterprise Strategy via Visual Analytics. *Communications of the ACM*, *59*(2), 26–28. https://doi.org/10.1145/2856447
Evans, P. C., & Gawer, A. (2016). The rise of the platform enterprise: a global survey. *The Center for Global Enterprise*, (January), 1–30.
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37.
Gawer, A., & Cusumano, M. A. (2008). How Companies Become Platform Leaders. *MIT Sloan Management Review*, *49*(2), 28. https://doi.org/Article
Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, *31*(3), 417–433. https://doi.org/10.1111/jpim.12105
Ghazawneh, A., & Henfridsson, O. (2011). Micro-strategizing in platform ecosystems: a multiple case study. In *Thirty Second International Conference on Information Systems*. Shanghai,.
Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal*, *23*(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x
Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611–642. Journal Article. https://doi.org/10.2307/25148742
Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A Taxonomy of Data-Driven Business Models Used by Start-up Firms*.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*. https://doi.org/10.1016/0010-0277(96)81418-1
Huang, P., Ceccagnoli, M., Forman, C., & Wu, D. J. (2009). Participation in a platform ecosystem: appropriability, competition and access to the installed base.
Iansiti, M., & Zhu, F. (2007). Dynamics of platform competition: Exploring the role of installed base, platform quality and consumer expectations. In *ICIS 2007 Proceedings* (p. 38). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871158327&partnerID=40&md5=c6dd685afc4ba3b3d973c896c0131996
Jagadish, H. V, Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86–94.
Jarvenpaa, S., & Markus, M. L. (2018). Genetic Platforms and Their Commercialization: Three Tales of Digital Entrepreneurship. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.
Katz, A., & Te’eni, D. (2007). The Contingent Impact of Contextualization on Computer-Mediated Collaboration. *Organization Science*. https://doi.org/10.1287/orsc.1060.0237
Kelestyn, B., & Henfridsson, O. (2014). Everyday digital entrepreneurship: The inception, shifts, and scaling of future shaping practices. In *Proceedings of the Thirty Fifth International Conference on Information Systems* (p. 8). Auckland, New Zealand.
Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 67–93.
Koutroumpis, P., & Leiponen, A. (2013). Understanding the value of (big) data. In *Proceedings of 2013 IEEE international conference on big data* (pp. 38–42).
Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature*, *470*(7333), 187.
Lee, P. (2015). Centralization, Fragmentation, and Replication in the Genomic Data Commons. In B. M. Frischmann, M. J. Madison, & K. J. Strandburg (Eds.), *Governing Medical Knowledge Commons* (pp. 46–73). Cambridge University Press. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2650189
Majchrzak, A., Malhotra, A., & John, R. (2005). Perceived individual collaboration know-how development through information technology-enabled contextualization: Evidence from

distributed teams. *Information Systems Research*, *16*(1), 9–27. https://doi.org/10.1287/isre.1050.0044

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big Data. *The Management Revolution. Harvard Bus Rev*, *90*(10), 61–67.

McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, *38*(1), 141–160.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., … Lopez, R. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research*, *41*(W1), W597–W600.

Nambisan, S. (2016). Digital Entrepreneurship: Toward a Digital Technology Perspective of Entrepreneurship. *Entrepreneurship Theory and Practice*. https://doi.org/10.1111/etap.12254

Nambisan, S., & Baron, R. A. (2013). Entrepreneurship in Innovation Ecosystems: Entrepreneurs' Self-Regulatory Processes and Their Implications for New Venture Success. *Entrepreneurship Theory and Practice*, *37*(5), 1071–1097. https://doi.org/10.1111/j.1540-6520.2012.00519.x

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *Mis Quarterly*, *41*(1).

OECD. (2015). *Digital Transformation of European Industry and Enterprises: : A Report of the Strategic Policy Forum on Digital Entrepreneurship*.

Oxley, J. E. (1999). Institutional environment and the mechanisms of governance: the impact of intellectual property protection on the structure of inter-firm alliances. *Journal of Economic Behavior & Organization*, *38*(3), 283–309.

Parker, G., Alstyne, M. Van, & Jiang, X. (2017). Platform Ecosystems: How Developers Invert the Firm. *Management Information Systems Quarterly*, *41*(1), 255–266.

Piezunka, H. (2011). Technological platforms. *Journal Für Betriebswirtschaft*, *61*(2), 179–226. https://doi.org/10.1007/s11301-011-0078-x

Pisano, G. P., & Teece, D. J. (2007). How to capture value from innovation: Shaping intellectual property and industry architecture. *California Management Review*, *50*(1), 278–296.

Project MinE. (2015). Beyond Project MinE NL. Retrieved November 23, 2018, from https://www.projectmine.com/beyond-project-mine-nl/

Rietveld, J., & Eggers, J. P. (2018). Demand Heterogeneity in Platform Markets: Implications for Complementors. *Organization Science*, *29*(2), 304–322.

Sanchez, R., & Mahoney, J. T. (1996). Modularity, flexibility, and knowledge management in product and organization design. *Strategic Management Journal*, *17*(S2), 63–76.

Silverman, D. (2016). *Qualitative research*. Sage.

Sussan, F., & Acs, Z. J. (2017). The digital entrepreneurial ecosystem. *Small Business Economics*, 1–19.

Te'eni, D. (2001). Review: A Cognitive-Affective Model of Organizational Communication for Designing IT. *MIS Quarterly*, *25*(2), 251. https://doi.org/10.2307/3250931

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305.

Teece, D. J. (1998). Capturing Value from Knowledge Assets. *California Management Review*, *40*(3). Retrieved from http://apps.business.ualberta.ca/mlounsbury/techcom/readings/teece.pdf

Tiwana, A. (2008). Does technological modularity substitute for control? A study of alliance performance in software outsourcing. *Strategic Management Journal*, *29*(7), 769–780.

Tiwana, A. (2014). The Rise of Platform Ecosystems. In *Platform Ecosystems* (pp. 3–21). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-408066-9.00001-1 M4 - Citavi

Tiwana, A. (2015). Evolutionary Competition in Platform Ecosystems. *Information Systems Research*, *26*(2), 266–281.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary-Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, *21*(4), 675–687.

Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2016). A Commons Perspective on Genetic Data Governance: the Case of Brca Data. In *ECIS* (p. ResearchPaper136).

Voelkerding, K. V, Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, *55*(4), 641–658.

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330.

Weick, K. E., & Meader, D. K. (1993). Sensemaking and group support systems. *Group Support Systems: New Perspectives*, 230–252.

West, J. (2017). Open Source Platforms Beyond Software: From ICT to Biotechnology. In *Entrepreneurship, Innovation, and Platforms* (pp. 337–370). Emerald Publishing Limited.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary-The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*(4), 724–735.

Yu, S., & Woodard, C. J. (2008). Innovation in the programmable web: Characterizing the mashup ecosystem. In *International Conference on Service-Oriented Computing* (pp. 136–147). Springer.

Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, *33*(1), 88–106. https://doi.org/10.1002/smj.941